What is claimed is:

1.   A character string recognition apparatus, comprising:

5   a key character code extraction unit automatically extracting a code string of a key word which is a node of a character string from a character string category to be recognized and expressed as a character code;

10   a key word extraction unit extracting a key word extracted by said key character code extraction unit or a part of the key word is extracted from a character string image; and

        a recognition unit holistically recognizing

15 character strings in partial areas determined by the extracted key words.

2.   The apparatus according to claim 1, further comprising

20   a verification unit verifying a recognition result of the holistic recognition by said recognition unit.

3.   The apparatus according to claim 1, wherein

25   when a key word is extracted from a character

string image, and when only a part of a character forming the key word is extracted, an extraction condition as a key character for preceding and subsequent characters is mitigated, and a character is re-extracted.

4.   The apparatus according to claim 1, wherein when a key word is extracted from a character string image, and when leading and trailing characters in the characters in the key word, and more than a predetermined ratio of the characters forming the key word are extracted, said key word extraction unit regards a partial character string as a key word.

5.   The apparatus according to claim 1, wherein when a key word is extracted from a character string image, when two or more separate characters are extracted in the characters forming the key word, and when more than a predetermined ratio of the characters in an area enclosed by the extracted characters, said key word extraction unit extracts the partial character string as a partial character string of the key word.

6.   The apparatus according to claim 1, wherein
when a key word is extracted from a character
string image, said key word extraction unit
performs a holistic recognizing process on an
extracted key word or a partial key word, and
verifies probability as a word.

7.   The apparatus according to claim 1, wherein
when a key word is extracted from a character
string image, said key word extraction unit
compares an area segmented as one character in
character feature and word feature, and extracts a
character string forming part of a key word or the
key word.

8.   The apparatus according to claim 1, wherein
when a word is extracted using word feature of
a key word from a character string image, said key
word extraction unit enhances recognition precision
in word recognition by referring to a dictionary in
which a word easily misrecognized as a key word is
entered as a similar word.

9.   The apparatus according to claim 1, wherein
when a code string of a key word which is a

node of a character string is extracted from a character string category, said key character code extraction unit extracts a character having a large number of occurrences in entire character strings to be recognized, a character having a large number of occurrences in a character string unit, and/or a set of closely associated characters as key words.

10.   The apparatus according to claim 1, wherein

    A character which is not easily misrecognized is entered in advance, and said key character code extraction unit extracts the entered character as a key character when a code string of a key word as a node of a character string from a character string category.

11.   The apparatus according to claim 1, wherein

    when a word area is holistically recognized, said recognition unit performs a word recognizing process, segments a character for the area, and recognizes the character so that a word recognition result can be determined when a character contained in the word recognition result is contained as n higher order and has a number of occurrences equal to or larger than a threshold in the character

recognition result.

12. The apparatus according to claim 2, wherein:

said recognition unit holistically recognizes a word area based on a word feature generated by combining character features;

said verification unit computes a division position of each character in a word image from a matching template, compares line density of a word image obtained at each division position with line density held by each character of a recognized word, and rejects a word recognition result when a sum of the line density, or a difference in a collation ratio is larger than a threshold.

13. The apparatus according to claim 2, wherein:

said recognition unit holistically recognizes a word area based on a word feature generated by combining character features;

said verification unit computes a division position of each character in a word image from a matching template, compares peripheral distribution of a word image obtained at each division position with peripheral distribution held by each character of a recognized word, and rejects a word

recognition result when a sum of the peripheral distribution, or a difference in a collation ratio is larger than a threshold.

5   14.   The apparatus according to claim 2, wherein:

said recognition unit holistically recognizes a word area based on a word feature generated by combining character features;

said verification unit compares a number of
10   characters in a recognized word is compared with a number of characters estimated from a word image, and rejects a word recognition result when a difference in the number of characters is larger than a threshold.

15

15.   A character string recognition apparatus, comprising:

key character code extraction means for automatically extracting a code string of a key
20   word which is a node of a character string from a character string category to be recognized and expressed as a character code;

key word extraction means for extracting a key word extracted by said key character code
25   extraction means or a part of the key word is

extracted from a character string image; and

recognition means for holistically recognizing character strings in partial areas determined by the extracted key words.

16. A character string recognition apparatus, comprising:

a recognition target character string group storage unit storing a list of character strings in a category to be recognized; and

a key word determination unit searching said recognition target character string group storage unit for each character to obtain a number of occurrences of each character, defining a character having a large number of occurrences as a key character, and defining a character string having a large number of occurrences as a key word.

17. A character string recognition apparatus, comprising:

a key character/word storage unit storing a determined key character or key word; and

a key character/word extraction unit extracting a character string as a key word if a part of the character string in the key word is

extracted when a key character or a key word stored in said key character/word storage unit is extracted from a character string to be recognized.

5   18.   A character string recognition apparatus, comprising:

a word recognition unit recognizing a word; and

a verification unit checking whether or not a

10 recognition result of said word recognition unit is correct.

19.   The apparatus according to claim 18, wherein said verification unit verifies a recognition

15 result based on line density or peripheral distribution.

20.   A key word determining method, comprising the step of

20 obtaining a number of occurrences of each character in a list stored in advance based on the list of character strings in a category to be recognized, defining a character having a large number of occurrences as a key character, and

25 defining a character string having a large number

of occurrences as a key word.

21. A   character   string   recognizing   method,
comprising the steps of:

5   obtaining a number of occurrences of each
character in a list stored in advance based on the
list of character strings in a category to be
recognized, defining a character having a large
number of occurrences as a key character, and
10   defining a character string having a large number
of occurrences as a key word;

     extracting the key character or the key word
from a character string image to be recognized; and

     recognizing a word for each area delimited by
15   each key character or key word in the character
string image to be recognized.

22.   A computer-readable storage medium storing a
program used to direct a computer to realize the
20   functions comprising

     obtaining a number of occurrences of each
character in a list stored in advance based on the
list of character strings in a category to be
recognized, defining a character having a large
25   number of occurrences as a key character, and

defining a character string having a large number of occurrences as a key word.

23. A computer data signal embodied in a carrier wave and representing a program that makes a computer to control interchanging data concerning a process included in a series of process flows with an external device, and the program making the computer execute the steps of:

obtaining a number of occurrences of each character in a list stored in advance based on the list of character strings in a category to be recognized, defining a character having a large number of occurrences as a key character, and defining a character string having a large number of occurrences as a key word;

extracting the key character or the key word from a character string image to be recognized; and

recognizing a word for each area delimited by each key character or key word in the character string image to be recognized.

24. A storage medium storing a program recognizing a character string image, said program comprising the processes of:

automatically extracting a code string of a key word which is a node of a character string from a character string category to be recognized and expressed as a character code;

5   extracting the extracted key word or a part of the key word from a character string image; and

holistically recognizing character strings in partial areas determined by the extracted key words.